DOCUMENT RESUME

ED 374 146 TM 022 067

AUTHOR Ingels, Stephen J.; Owings, Jeffrey A.
TITLE Methodological Issues Encountered in Following a Cohort of Eighth Graders.
PUB DATE 26 May 94
NOTE 20p.
PUB TYPE Reports - Research/Technical (143)

EDRS PRICE MF01/PC01 Plus Postage.
DESCRIPTORS *Attrition (Research Studies); Change; *Cohort Analysis; *Eligibility; Grade 8; High Schools; High School Students; Junior High Schools; *Junior High School Students; Longitudinal Studies; National Surveys; Research Design; *Research Methodology; Research Problems; Sample Size; *Sampling
IDENTIFIERS *National Education Longitudinal Study 1988; Nonresponders; Transitional Activities

ABSTRACT

The National Education Longitudinal Study of 1988 (NELS:88) serves as an example of how three specific problems of representativeness in a longitudinal study may be approached and overcome. NELS:88 was designed to provide longitudinal data about the educational transitions experienced by students as they leave eighth grade and move through high school and postsecondary education. The cohort, initially 26,432 students, is being followed at 2-year intervals. The three key issues that are discussed are: (1) eligibility and exclusion rules and the measures taken to deal with undercoverage resulting from exclusion; (2) the need for sample freshening to ensure representative sophomore and senior cohorts in 1990 and 1992; and (3) attempts to minimize sample attrition and nonresponse error. Longitudinal designs are a powerful vehicle for measuring individual-level change. At the same time, they have distinct limitations, which can be overcome by careful execution and research design. One figure and three tables illustrate the discussion. (Contains 21 references.) (SLD)

****
* Reproductions supplied by EDRS are the best that can be made *
* from the original document. *
****

ED 374 146

U.S. DEPARTMENT OF EDUCATION
Office of Educational Research and Improvement
EDUCATIONAL RESOURCES INFORMATION CENTER (ERIC)
☒ This document has been reproduced as received from the person or organization originating it.
☐ Minor changes have been made to improve reproduction quality.
• Points of view or opinions stated in this document do not necessarily represent official OERI position or policy

"PERMISSION TO REPRODUCE THIS MATERIAL HAS BEEN GRANTED BY
STEVEN J. INGELS
TO THE EDUCATIONAL RESOURCES INFORMATION CENTER (ERIC)"

Council of Professional Associations on Federal Statistics

SEMINAR ON NEW DIRECTIONS IN STATISTICAL METHODOLOGY
Session on Longitudinal Surveys

Methodological Issues Encountered in Following a Cohort of Eighth Graders

Steven J. Ingels, National Opinion Research Center (NORC) at the University of Chicago, and
Jeffrey A. Owings, National Center for Education Statistics (NCES), U.S. Department of Education.

Bethesda, Maryland
May 26, 1994

This paper is based on the National Education Longitudinal Study of 1988 (NELS:88). NELS:88 is a longitudinal micro-database sponsored by the National Center for Education Statistics, U.S. Department of Education. The ideas and opinions expressed in this paper are those of the authors and do not necessarily represent positions or conclusions endorsed by NCES.

Methodological Issues Encountered in Following a Cohort of Eighth Graders

Steven J. Ingels, Jeffrey A. Owings

Introduction. Longitudinal studies typically employ a probability sample of a unit (for example, individuals, institutions [e.g., schools], groups [e.g., families] or inanimate objects [e.g., dwelling units]) that is drawn at one point in time, then repeatedly observed, so that change in units can be measured over time. NELS:88 gathers longitudinal data from a panel of students (including those who exit school prior to completion); institutional data are collected from three cross-sections of schools that are associated with the panel of students.[1]

Longitudinal designs provide a powerful vehicle for reliably[2] measuring individual-level change and development as well as for describing the dynamics of change and the processes that are associated with it. At the same time, longitudinal studies have both inherent and potential limitations (see, for example, Pearson 1989 for a discussion of advantages and disadvantages of longitudinal surveys).

Some of the most important of these limitations can in large measure be overcome if the study is properly executed or if potential limitations are explicitly addressed in the design. This is the case, in particular, for three specific threats to longitudinal sample representativeness.[3]

(1) *Undercoverage*. Potentially biasing undercoverage[4] may arise from any of several sources. It may arise from deliberate or inadvertent exclusion of part of the baseline "target" sample, or may arise if baseline nonrespondents are not pursued in subsequent waves. If follow-up rounds are to be representative of some additional populations (as is the case in NELS:88 which, through sample freshening (see [2] below) generates representative sophomore and senior cohorts in the 1990 and 1992 follow-ups) undercoverage problems may also arise if the eligibility of ineligible baseline students whose eligibility status is subject to change is not reassessed in succeeding rounds.

(2) *Need for Freshening*. The unit or cohort being studied may become less statistically representative of the target population (or less policy-relevant) over time. For example, a sample of individuals in a given geographical area may become less representative of that area as sample members disperse and other individuals move into the area. A sample of eighth graders two years later is not fully representative of the nation's tenth graders at the second point in time.

(3) *Attrition and Nonresponse*. Sample attrition poses substantial risks for a longitudinal study's representativeness. This danger can be overcome if high response rates are maintained across all rounds, and may be partially compensated for in weighting.

In this paper, the National Education Longitudinal Study of 1988 (NELS:88) serves as an example of how these three specific problems of representativeness in a longitudinal study may be approached and overcome. We more briefly comment on some additional potential sources of survey error.

1. Description of NELS:88

As a point of entry into our topic, we briefly describe NELS:88 by summarizing its goals, surveys administered, response rates, and analysis potential.

1.1 *Goals of NELS:88*

Beginning in 1988 with a cohort of 26,432 eighth graders attending 1,052 public and private schools across the nation, NELS:88 was

designed to provide longitudinal data about critical transitions experienced by students as they leave eighth grade school settings, progress through high school (or drop out), enter and leave postsecondary institutions, and enter the work force. The 1988 eighth grade cohort has been followed at two-year intervals (specifically, first follow-up -- 1990; second follow-up -- 1992) with a third follow-up currently (spring 1994) underway, and a fourth follow-up to take place in 1997 or 1998.

Major features of NELS:88 include:

- the integration of student, dropout, parent, teacher, school administrator and school records (transcript) surveys;
- the initial concentration on an eighth grade student cohort with follow-ups at two year intervals;
- the inclusion of supplementary components to support analyses of geographically or demographically distinct subgroups (for example, selected state supplements; oversamples of Asians and Hispanics, and of students in private schools); and
- the design linkages to previous longitudinal studies (High School and Beyond [HS&B], the National Longitudinal Study of the High School Class of 1972 [NLS-72]) and other current studies (for example, the National Assessment of Educational Progress [NAEP] testing program and high school transcript data collections).

The longitudinal design of NELS:88 permits the examination of change in young people's lives and the role of schools, teachers, community, and family in promoting growth and positive life outcomes. In particular, data from NELS:88 can be used to investigate issues in the context of the family, community, school, and classroom including:

- Students' academic growth over time;
- The transition from eighth grade to high school and the transition from high school to the labor market or postsecondary education;
- The process of dropping out of school, as it occurs from the end of eighth grade on;
- The role of schools in helping the disadvantaged;
- The school experiences and academic performance of language minority students;
- The influence of ability grouping, program type, and coursetaking patterns on future educational persistence and achievement; and
- The features of effective schools.

1.2 Surveys administered

NELS:88 components, by wave, are summarized in Figure 1.

Figure 1: Base Year Through Fourth Follow-Up -- NELS:88 Components

BASE YEAR	FIRST FOLLOW-UP	SECOND FOLLOW-UP	THIRD FOLLOW-UP	FOURTH FOLLOW-UP
spring term 1988 GRADE 8	spring term 1990 MODAL GRADE = SOPHOMORE	spring term 1992 MODAL GRADE = SENIOR	spring 1994 MODE = H.S. + 2 YEARS	spring 1997 (or 1998) HS + 5 YEARS (or 6)
Students: Questionnaire, Tests*	**Dropouts, Students:** Questionnaire, Tests	**Dropouts, Students:** Questionnaire, Tests, H.S. Transcripts	**All Individuals:** Questionnaire	**All Individuals:** Questionnaire
Parents: Questionnaire		**Parents:** Questionnaire		
Principals: Questionnaire	**Principals:** Questionnaire	**Principals:** Questionnaire		
Two Teachers per student: (taken from English, social studies, mathematics, science)	**Two Teachers per student:** (taken from English, social studie mathematics, or science)	**One Teacher per student:** (taken from mathematics or science)		

* Reading, social studies, math and science tests are administered in the three in-school rounds.

1.3 *Analysis potential - longitudinal vs. cross-sectional applications.*

Analytic Levels. The NELS:88 design enables researchers to conduct analyses on three principal levels: (1) within-wave (or cross-sectional) analysis at a single time point, (2) cross-cohort analysis (by comparing cross-sectional NELS:88 findings to those of comparable populations studied earlier in HS&B and NLS-72) and (3) cross-wave (or longitudinal[5]) analysis.

The first analytic level within NELS:88 is cross-sectional. By beginning with a cross-section of 1988 eighth graders, following a substantial subsample of these students at two-year intervals, and freshening the 1990 and 1992 samples to obtain representative national cross-sections of tenth and twelfth graders, the study also provides a statistical profile of America's eighth graders, high school sophomores, and high school seniors.

A second analytic level extends representative cross-sections to intercohort comparisons. NELS:88 provides researchers with data for drawing comparisons with previous NCES longitudinal studies. After the release of NELS:88 first follow-up data, researchers were able to conduct trend analyses with the 1980 sophomore cohort of HS&B. With completion of the NELS:88 second follow-up, comparisons may be made among NELS:88, HS&B, and NLS-72 senior cohorts. To facilitate cross-cohort comparisons, some of the questionnaire items used in the NLS-72 and HS&B high school surveys were repeated in NELS:88, and data processing and file conventions were kept consistent, to the maximum possible extent, with HS&B and NLS-72.[6]

The third analytic level is longitudinal, and utilizes repeated measurements on the same individuals over time. However, because NELS:88 comprises three nationally representative grade- and year-defined cross-sections, it supports multiple panels:

1988	eighth graders two, four, six, and nine years later
1990	sophomores two, four, and seven years later
1992	seniors two and five years later[7]

Change Analysis. Cross-sectional analysis provides a snapshot at a single point in time. Repeated cross-sectional analysis, and longitudinal analysis, permit the measurement of change over time. Change (and stability) over time can be measured at the group or individual level:

(1) *At the group level*, change can be measured across the successive cross-sections--eighth graders in 1988, sophomores in 1990, and seniors in 1992. In the same way, multicohort assessments such as NAEP can estimate overall and subgroup gains in specific subject matter proficiency across selected points in the school career (e.g., between fourth, eighth, and twelfth grade). In addition, NELS:88 and comparable studies (e.g., NLS-72 and HS&B) can be analyzed as repeated cross-sections. For example, 1972, 1980/82, and 1992 seniors can be compared, and trends measured, including changes in transition rates (e.g., trends in postsecondary access for various subpopulations). A cross-sectional time-series such as NAEP also measures trends (e.g. in math achievement for 17 year olds from 1973 to 1990 for the nation and subgroups).

A principal weakness of change measurement at the group level--whether one is looking at rolling (e.g., eighth graders in 1988, sophomores in 1990, seniors in 1992) or repeated (e.g., eighth graders in 1988, 1990, and 1992) static cross-sections[8] is that it sometimes masks individual change; high levels of individual change are not incompatible with stability at the aggregate level. Thus, for example, looking at the proportion of 1988 eighth graders in 1988 who were out of school

in 1990 (6.8%) and comparing this to the proportion out of school in 1992 (11.6%) masks the cumulative number of individuals who were 1990 or 1992 spring term dropouts, since some 1990 dropouts had returned to school by 1992.[9] A *locus classicus* of this phenomenon is found in studies of poverty and welfare recipience. While the proportion of adolescent mothers receiving AFDC over time is relatively constant, the AFDC population is not. Mobility onto, and off, the AFDC's rolls is demonstrated by longitudinal data provided by the NLSY, but would not be apparent from repeated cross-sectional results. Likewise, PSID data show that while poverty rates may be roughly stable over time, poverty spells for individuals and households tend to be relatively brief.[10]

(2) Change can also be analyzed *at the individual level* over time. The latter possibility--true longitudinal measurement in which precursor events can be related to outcomes--represents, for most purposes, the unique strength of the NELS:88 design. Following individual educational histories generally provides the best basis for drawing causal inferences about educational processes and their effects. Two broad kinds of analysis scenarios are possible. Longitudinal analysis can involve repeated measures of the same outcome--for example, test data can be used to measure growth in academic achievement over time. Or longitudinal analysis can show how conditions at an earlier time point are predictive of outcomes at a later time point. For example, one might examine how eighth graders with single or clustered "risk factors" (for example, such status risk factors as coming from a low-income home, having parents who did not finish high school, and so on; or such behavioral risk factors as cutting classes, lack of participation in extracurricular activities, and so on) fared two years later (for example, what proportion had dropped out, repeated a grade, and so on).

While longitudinal studies are prospective, in that they offer the opportunity to record new events, longitudinal analysis may be either retrospective or prospective. In NELS:88, priority in the baseline was given to questions predictive of future behavior. However, while questions that asked for *reasons* for past behavior were deliberately avoided, some retrospective questions were posed, when their focus was on simple descriptions of salient past events. For example, parents were asked whether their eighth grader had attended a Head Start program[11] or kindergarten or preschool, whether other of their children (respondent's elder siblings) had dropped out of school, and so on.

Multilevel Analyses. Educational research often deals with nested hierarchies, such as students within classrooms, students and classrooms within schoools, schools within districts, and so on. Longitudinal analysis can be a powerful device for measuring school and program effectivness or disentangling the effects of school and home. However, in addition to sample attrition, sample *dispersion* poses a threat to multilevel longitudinal analysis. This problem poses some difficulty within the context of following, say, a high school sophomore cohort through senior year (as in HS&B), but poses enormous difficulty for measuring high school effects when one begins with a cohort of eighth graders. By the time a cohort of eighth graders reaches high school, in-school clusters are much diminished in size, are no longer necessarily representative of students in the school, nor do the high schools to which eighth graders have dispersed constitute a national probability sample of high schools. Hence, while one may use follow-up data to continue to trace the longer-term individual-level effects of different kinds of eighth grade schools, effects of secondary schools can only be studied if extraordinary extra measures are taken. These include augmenting the student sample within schools to make it larger and representative, and developing estimators so that selection probabilities can be approximated for high schools.[12] These extraordinary measures have been implemented in a subsample of NELS:88 high schools to facilitate school effectiveness

research.

2. Sample Representativeness

This section discusses three key issues. *First*, eligibility and exclusion rules, particularly as applied in the NELS:88 base year, and the measures taken in later rounds of the study to deal with the potential for undercoverage biases that might result from these exclusions. *Second*, the need for sample freshening to ensure representative sophomore and senior cohorts in 1990 and 1992, and the procedures undertaken to bring that sample freshening about. And *third*, attempts to minimize sample attrition and nonresponse error.

2.1 Eligibility: Excluded Students and Undercoverage Bias.

In the base year of NELS:88, students were sampled through a two-stage process. First, stratified random sampling and school contacting resulted in the identification of the school sample; second, students were randomly selected (with oversampling of Hispanics and Asians) from within cooperating schools.

The target population for the base year comprised all public and private schools containing eighth grades in the fifty states and the District of Columbia. Excluded from the NELS:88 school sample are Bureau of Indian Affairs (BIA) schools, special education schools for students with disabilities, area vocational schools that do not enroll students directly, and schools for dependents of U.S. personnel overseas.[13] The student population excludes students with severe mental disabilities, students whose command of the English language was not sufficient for understanding the survey materials (especially the cognitive tests), and students with physical or emotional problems that would make it unduly difficult for them to participate in the survey. This paper discusses (1) the consequences of student exclusion for the research design and results, and (2) the special measures that have been undertaken in NELS:88 to compensate or correct for the effects of exclusion. Before either of these two topics is pursued in detail, however, it will be desirable to say more about student exclusion in the NELS:88 base year--the 1987-88 school year during which the eighth grade cohort was selected and surveyed.

To better understand how excluding students with mental disabilities, language barriers, and severe physical and emotional problems affects population inferences, data were obtained on the numbers of students excluded as a result of these restrictions.

Seven ineligibility codes defining categories of excluded students were employed at the time of student sample selection:

A - attended sampled school only on a part-time basis, primary enrollment at another school.

B - physical disability precluded student from filling out questionnaires and taking tests.

C - mental disability precluded student from filling out questionnaires and taking tests.

D - dropout: absent or truant for 20 consecutive days, and was not expected to return to school.

E - did not have English as the mother tongue AND had insufficient command of English to complete the NELS:88 questionnaires and tests.

F - transferred out of the school since roster was compiled.

G - was deceased.

Before sampling, school coordinators--members of the school staff, typically an assistant principal or guidance counselor who acted as liaison between the school and the study--were asked to examine the school sampling roster and annotate each excluded student's entry by assigning one of the exclusion codes. Because eligibility decisions were to be made on an individual basis, special education and Limited English Proficiency (LEP) students were not to be excluded categorically. Rather, each student's case was to be reviewed to determine the extent of limitation in relation to the prospect for meaningful survey participation. Each individual student, including LEPs and physically or mentally handicapped students, was to be designated eligible for the survey if school staff deemed the student capable of completing the NELS:88 instruments, and excluded if school staff judged the student to be incapable of doing so. School coordinators were told that when there was doubt, they should consider the student capable of participation in the survey. Exclusion of students after sampling ("post-roster ineligibles") occurred either during the sample update just prior to survey day, or on survey day itself. Such exclusion after sampling normally occurred because of a change in student status (for example, transfer, death). However, in very rare instances such exclusions reflected belated recognition of a student's pre-existing ineligibility--that is, if an annotation error was made and an ineligible student selected for the sample in consequence of such an error, ineligibility became apparent later in the survey, whereupon the student was excluded.

Excluded students were divided into those who were full-time students at the school (categories B, C, and E) and those who were not (categories A, D, F, & G). Our main concern here is with students who were full-time students at the school but who were excluded from the sample. Excluding these students will affect estimates made from the sample.

Students in categories A (n=329), D (n=733), F (n=3,325), and G (n=6) were either not at the school or were present only part time (with primary registration at another school, hence a chance of selection into NELS:88 at another school). Thus excluding students in these categories has no implications for making estimates to the population of eighth grade students.

It should be noted that students in category F, those who had transferred out of the sampled school, had some chance of being selected into the sample if they transferred into another NELS:88 sampled school just as transfers into NELS:88 schools from non-NELS:88 schools had a chance of selection at the time of the sample update. The sampling of transfer-in students associated with the sample update allowed NORC to represent transfer students in the NELS:88 sample.

The total eighth grade enrollment for the NELS:88 sample of schools was 202,996. Of these students, 10,853 were excluded owing to limitations in their language proficiency or to mental or physical disabilities. Thus 5.37 percent of the potential student sample (the students enrolled in the eighth grade in the 1,052 NELS:88 schools from which usable student data were obtained) were excluded. Less than one half of one percent of the potential sample was excluded for reasons of physical or emotional disability (.41 percent), but 3.04 percent was excluded for reasons of mental disability, and 1.90 percent because of limitations in English proficiency. (Note that these are raw sample percents, not weighted population estimates.)

Put another way, of the 10,853 excluded students, about 57 percent were excluded for mental disability, about 35 percent owing to language problems, and less than 8 percent because of physical or emotional disabilities. Because current characteristics and probable future educational outcomes for these groups may depart from the national norm, the exclusion factor should be taken into

consideration in generalizing from the NELS:88 sample to eighth graders in the nation as a whole. This implication for estimation carries to future waves. For example, if the overall propensity to drop out between the eighth and tenth grades is twice as high for excluded students as for non-excluded students, the dropout figures derivable from the NELS:88 first follow-up (1990) study would underestimate early dropouts by about ten percent. (In point of fact, the 1988-90 status dropout rate derivable from the eligible NELS:88 sample representing about 94.6 percent of the cohort is between 6.0 and 6.1 percent, and from the expanded--eligible + ineligible--1988 sample representing [virtually] 100 percent of the cohort, 6.8 percent.)

Undercoverage of course affects the power of a study both to produce national estimates, and, yet more dramatically, to produce estimates for the particular group that is not fully covered.[14] Undercoverage, moreover, poses some special difficulties for the representativeness of a multi-cohort longitudinal study such as NELS:88.

In a school-based longitudinal survey such as NELS:88, baseline excluded students affect the representativeness of freshened grade cohorts in future waves. To achieve a thoroughly representative tenth grade (1990) and twelfth grade (1992) sample comparable to the High School and Beyond 1980 sophomore cohort (or, for 1992, the HS&B 1980 senior cohort and the base year of NLS-72), the NELS.88 follow-up samples must approximate those which would have come into being had a new baseline sample independently been drawn at either of the later time points. In 1990 (and 1992) one must therefore freshen, to give "out of sequence" students (for example, in 1990, those tenth graders who were not in eighth grade in the spring of 1988) a chance of selection into the study. One must also accommodate excluded students whose eligibility status has changed, for they too (with the exception of those who fell out of sequence in the progression through grades) would potentially have been selected had a sample been independently drawn two years later, and must have a chance of selection if the representativeness and cross-cohort comparability of the follow-up sample is to be maintained. Thus, for example, if a base year student excluded because of a language barrier achieves the level of proficiency in English that is required for completing the NELS:88 instruments in 1990 or 1992, that student should have some chance of re-entering the sample.

A substantial subsample of the base year ineligibles was, accordingly, followed in 1990 and 1992, to reassess eligibility status and gather information about excluded students' demographic characteristics, educational paths, and life outcomes. Data on persistence in school to be obtained from this subsample has been used to derive an adjustment factor for national estimates of the eighth grade cohort's dropout rates between spring of 1988 and spring of 1990, and from 1988 and 1990 to 1992.

The base year ineligibles study largely compensates for population undercoverage. Small populations who remain outside the baseline sampling frame include students who are educated at home or in private tutorial settings, those who are in excluded categories of schools[15] and those who have dropped out of school before reaching the eighth grade.

Table 1 shows that by 1992, a substantial portion of the sample of base year ineligible students had been reclassified as eligible. Excluded students who were later classified as eligible were included in NELS:88 follow-up surveys.

Reclassifications reported in Table 1 reflect multiple phenomena. In some cases--and presumably this is particularly the case for the language exclusions--reclassification reflects change in the eligibility status of the sample member over time. In other cases, change represents the unreliability of exclusion judgments, particularly for exclusion reasons

that are more open to interpretation (e.g., mental as opposed to physical disabilities) or that apply to individuals at the margin of the classification--different individuals were asked to assess eligibility at different points in time. Finally, some of the change registered in Table 1 reflects the fact that in the follow-ups we provided more detailed interpretation for the guidelines, so that the validity of exclusion judgments would be enhanced. All in all, however, if any individuals in the target population are to be subject to exclusion from the baseline of a longitudinal study, it is of some importance to reassess their eligibility over time, particularly, in a school-based survey, if the panel is to represent additional grade cohorts.

Table 1: 1992 Status Ns of 1988 Excluded Students

1988 reason for exclusion:	ELIG.	INELIG.	OUT OF SCOPE	N.A.	SAMPLING ERROR
language	125	22	25	30	23
physical	13	9	0	1	1
mental	166	140	5	25	16
unknown	30	15	2	10	16
TOTAL	334	186	32	66	56

* N.A. = status not ascertained.

2.2 Representativeness and New Grade Cohorts: Sample Freshening.

Pearson (1989) notes that a potential limitation of longitudinal samples is that they may provide estimates of the population from which they were originally drawn, but not of the current population. It is of interest to follow a sample of 1988 eighth graders. Nevertheless, an eighth grade panel two years later will not by itself provide a representative sample of the nation's high school sophomores, nor four years later a representative sample of seniors. Representative sophomore and senior samples are analytically desirable at all three levels of NELS:88 analysis *First*, it is desirable to be able to make cross-sectional generalizations about the nation's sophomores in 1990 and seniors in 1992. *Second*, it is desirable to be able to make intercohort comparisons between

HS&B 1980 sophomores and 1990 NELS:88 sophomores; between NLS-72 (1972) and HS&B (1980) seniors and NELS:88 (1992) seniors; and between the transcript records of HS&B (1982), NAEP (1987 and 1990), and NELS:88 (1992) seniors. *Third*, it is desirable to be able to conduct longitudinal analyses of 1990 sophomores two, four, and more years later, and of 1992 seniors two and more years later.

Hence a major sampling objective of NELS:88 was to create a valid probability sample of students enrolled in tenth grade in the spring term of the 1989-1990 school year and of students enrolled in the twelfth grade in the spring term of the 1991-92 school year. This goal was achieved by a process we have termed "freshening." The 1990 freshening procedure was carried out in four steps:

1. For each school that contained at least one base year 10th grade student selected for interview in 1990, a complete alphabetical roster of all 10th grade students was obtained.

2. For each base year sample member, we examined the next student on the list; if the base year student was the last one listed on the roster, we examined the first student on the roster (that is, the roster was "circularized").

3. If the student who was examined was enrolled in the 8th grade in the U.S. in 1988, then the freshening process terminated. If the designated student was not enrolled in the 8th grade in the U.S. in 1988, then that student was selected into the freshened sample.

4. Whenever a student was added to the freshened sample in step 3, the next student on the roster was examined and step 3 was repeated. The sequence of steps 3 and 4 was repeated (adding more students to the freshened sample) until a student who was in the 8th grade in the U.S. in 1988 was reached on the roster.

At a given first follow-up school, the freshening process could yield zero, one, or more than one new sample member. Altogether, 1,229 new students were added to the tenth grade sample--on average, just less than one student per school.[16] This procedure was repeated in 1992, to generate a probability sample of the nation's high school seniors.

This freshening procedure is an essentially unbiased method[17] for producing a probability sample of students who were enrolled in the tenth grade in 1990 (or twelfth grade in 1992) but were not enrolled in the eighth grade in the U.S. in 1988. There is a very small bias introduced by the omission of eligible tenth (or twelfth) graders attending schools that included *no* students who were eighth graders in 1988. There is an additional small bias introduced by not freshening on the members of the sample of base year ineligibles. All other 1990 sophomores (or 1992 seniors) who qualify for the freshening sample have some chance of selection. This is because every student who was in the tenth grade in 1990 (or twelfth grade in 1992) but not in the eighth grade in 1988 is linked to exactly one student who was a 1988 eighth grader--this is the 1988 eighth grader who would immediately precede the candidate for the freshening sample on a circularized, alphabetical roster of tenth graders at the school. Because each 1988 eighth grader had a calculable, non-zero probability of selection into the base year and first follow-up samples, we can calculate the selection probabilities for all students eligible for the freshening sample. Thus, the freshening procedure produces a student sample that meets the criterion for a probability sample.

As noted above (1.3), the NELS:88 school sample in 1990 and 1992--the schools to which 1988 eighth graders matriculated--was not nationally representative. Hence for a select subset of schools, in order to provide a basis for studying school effects, feeder pattern information was collected so that tenth grade

school selection probabilities could be approximated, and student samples augmented to make them robust and representative of the school's tenth grade class.[18]

2.3 Nonresponse Error as a Potential Source of Bias: Measures to Maximize Response Rates.

Cumulative nonresponse poses a special threat to longitudinal studies. Some individuals are missed in the baseline measurement, and may enter the study late. Other individuals may be lost, through mobility and the inability to locate them at a later date, or may cease to participate in the study. Still others may participate in the baseline, become temporarily out of scope by leaving the country or become nonrespondents by refusing to participate in the initial follow-up, then re-enter the study in a later follow-up. A longitudinal study must maximize the number of individuals who have data at all data points. Although weighting may help to adjust for nonresponse, the representativeness of the panel depends, in the final analysis, on maintaining high participation rates.

NELS:88 Response Rates. High response rates have been achieved by the study. In the NELS:88 base year (1988) 93.1 percent of selected eighth graders participated. In the NELS:88 first follow-up (1990), 93.9 percent of student and dropout sample members (19,264 of 20,524) took part. In the second follow-up, 90.7 percent of student and dropout sample members took part.

However, from the point of view of longitudinal analysis, a more critical statistic is the proportion of the sample with data at all time points (or, the proportion of baseline participants with data for all follow-ups). Of the 18,261 base year participants retained in the first follow-up, 17,424--or 95.4 percent--were successfully resurveyed. From this base of eighth grade cohort members with both (1988 and 1990) data points, 95.1 percent were resurveyed in the second follow-up.

Table 2A shows overall and subgroup results for the base year-first follow-up respondents for whom a reinterview was attempted in 1992. While, as noted above, around 95 percent were successfully resurveyed (that is, completed a student or dropout questionnaire) in 1992 and thus have data for all three waves, far fewer (72 percent) completed the cognitive test in all three rounds. Table 2B depicts the across-round questionnaire completion status of base year-first follow-up participants who were second follow-up *students*, and the likelihood that school contextual data was available for them for all three rounds. These tables show that completion rates were very similar across different school control types, urbanicity, region, and high and low minority enrollment, and that similar response rates were obtained for members of different racial and ethnic groups.

However, even with these high rates of success in baseline and follow-up data collection, the proportion of the 1988 eighth grade cohort in 1992 with all three data points drops to 84 percent (16,489 of 19,645) when all students missing one or more data point owing to base year, first or second follow-up nonresponse or any other source of sample attrition--being deceased, sample members who suffered grave impairments in the course of the study that did not permit them to be surveyed, individuals out of scope for either follow-up round by virtue of being outside the country--are factored in.

Overall, then, NELS:88 has achieved reasonably high student panel response rates. In addition, final weights have been adjusted for nonresponse, using nonresponse adjustment cells based upon combinations of classification values reflecting race, gender, and data collection status (e.g., dropout; in school in expected grade; in school in another grade; and so on).[19]

Means of Achieving High Response Rates. The means by which these high response rates were achieved may be concisely summarized. Most individuals changed

schools, and many changed home addresses, between the base year and the follow-up surveys. About 99 percent of students were successfully traced between the base year and first follow-up, whereupon clusters of students were subsampled to reduce, for cost reasons, the number of high schools to be included in the study. The ability to successfully trace individuals was based upon extensive locating information collected in the base year from both students and parents. This locating information included name, address and telephone number for the student, each parent, and the family's closest relative or friend who did not live in the household. Eighth grade students were also asked to indicate what school they expected to be attending two years later. Tracing was carried out at two levels: first, it was ascertained if the sample member was at the expected school. If not, household information was used to locate the individual. In order to find base year nonrespondents (about 7 percent of the sample did not complete a 1988 student questionnaire and hence did not provide locating information), in addition to conventional survey locating sources, information about the schools matriculated to by the eighth grader's classmates was also utilized. Tracing procedures were repeated in the second follow-up, though between tenth and twelfth grade there is less dispersion to new schools and it was not necessary to further subsample students.

In order to survey students, contractor (NORC) staff administered the survey forms at a date agreeable to the school. Make-up sessions were conducted for students who missed the initial survey session. Dropouts and chronic absentees were pursued outside school. Such individuals were invited to group sessions and provided reimbursement for their travel expenses, or were interviewed in their households, over the telephone or in person.

In rare instances, NELS:88 has made use of respondent fees. For example, some dropouts received a monetary incentive, as did some high burden teachers (teachers who had to rate an unusually high number of NELS:88 students such that their burden of questionnaire completion might be two hours or more). School coordinators were given a modest honorarium (normally $25) for assisting with survey activities (for example, supplying annotated rosters, arranging space, and so on), but neither schools nor students were ever paid for their participation

TABLE 2A

NELS:88 Second Follow-Up data collection results for Base Year– First Follow-Up panel participants

	Student/Dropout questionnaire (BY, F1 and F2) Completion rates		Student/Dropout cognitive test[a] (BY, F1 and F2) Completion rates	
	Weighted	Unweighted	Weighted	Unweighted
Total				
Participated	94.7	95.1	69.6	72.2
Selected	16,489[b]		11,902	
School type[c]	17,337		16,489	
Public				
Catholic	94.3	94.7	69.0	71.4
Other private	97.9	97.0	74.1	78.6
Urbanicity[c]	97.4	97.0	73.0	73.7
Urban				
Suburban	93.5	95.1	64.3	69.5
Rural	95.5	95.3	69.1	70.1
Region[c]	94.8	94.9	74.6	77.2
Northeast				
South	94.8	95.1	70.3	71.3
Midwest	94.1	94.5	68.2	73.1
West	95.7	96.0	74.9	76.4
Ethnicity	94.6	95.1	63.7	65.7
Asian/PI				
Hispanic	93.3	95.0	71.5	71.9
Black	93.1	94.4	63.9	65.5
White	92.4	92.6	59.6	67.0
Am. Indian	95.5	95.7	72.1	74.2
Refused/Missing[d]	94.1	91.3	64.8	64.0
Minority schools[c]	81.1	75.0	38.3	55.6
Schools with more than 19% minority students				
Schools with less than 19% minority students	92.2	93.5	55.1	59.3
	95.0	95.3	71.0	73.5

[a] Cognitive test coverage rate for each sample member who has completed a BY student questionnaire, F1 and F2 student/dropout questionnaire.

[b] Sample members who participated in the BY, F1 and F2.

[c] Refers to 8th grade schools.

[d] Refused/Missing refers only to the status of a sample member's ethnicity. It does not refer to student/dropout nonparticipants.

TABLE 2B

NELS:88 Second Follow-Up student survey results for Base Year -- First Follow-Up panel participants

	Student questionnaire (BY, F1 and F2) Completion rates		School questionnaire[a] (BY, F1 and F2) Completion rates	
	Weighted	Unweighted	Weighted	Unweighted
Total	95.7	96.1	95.5	95.6
Participated	14,674[b]		13,182	
Selected	15,269		13,783	
School type[c]				
Public	95.4	95.8	95.8	95.7
Catholic	98.2	97.3	94.3	94.8
Other private	97.5	97.1	93.5	95.8
Urbanicity[c]				
Urban	94.4	96.4	93.7	94.7
Suburban	96.2	96.1	94.4	94.3
Rural	95.8	95.9	98.4	98.2
Region[c]				
Northeast	95.2	95.5	94.9	94.6
South	95.8	96.2	95.6	95.9
Midwest	96.2	96.5	97.5	97.8
West	95.5	96.0	93.1	93.2
Ethnicity				
Asian/PI	94.9	95.8	90.2	93.9
Hispanic	94.2	95.8	89.8	91.3
Black	94.3	95.0	95.1	95.3
White	96.2	96.4	96.5	96.5
Am. Indian	93.8	90.9	97.6	97.3
Refused/Missing[d]	74.2	72.7	100.0	100.0
Minority schools[c]				
Schools with more than 19% minority students	92.5	96.3	90.7	90.0
Schools with less than 19% minority students	96.0	94.4	96.0	96.2

[a] School questionnaire coverage rate for each student who completed a BY, F1, and F2 student questionnaire.

[b] Panel *students* only.

[c] Refers to 8th grade schools.

[d] Refused/Missing refers only to the status of a sample member's ethnicity. It does not refer to student nonparticipants.

3. Other Sources of Survey Error

When all is said and done, it is the total variable error and bias of a survey estimate that is critical (see Kish, 1965; Andersen, Kasper, and Frankel, 1979; Groves, 1989). From the point of view of total survey error, our discussion thus far is incomplete. It may be useful to identify additional sources of survey error, though space limitations do not permit us to address them.

There are various "repeated measurement" problems in longitudinal surveys. One of these problems is that of panel effects.[20] We do not believe that problems associated with repeated measurements (such as remembering past responses to individual items) are likely to be a difficulty, both because of the sheer number of test and questionnaire items asked, and the two year intervals between data collections. However, participation in a longitudinal study in theory may influence the survey member's subsequent behavior or attitudes.

There are many sources of measurement error. The validity of responses to the NELS:88 eighth grade questionnaire items has been examined in Kaufman, Rasinski, Lee and West (1991), which compares parent and student reports. Transcript and student reports were compared for the HS&B data by Fetters, Stowe and Owings (1984). Psychometric issues in the base year tests are addressed in Rock and Pollack (1991) and in a forthcoming second follow-up psychometric report.

Our earlier discussion dealt with unit nonresponse as a problem of maintaining individual participation across rounds. However, school nonresponse in the base year, and item nonresponse across the survey instruments, also are important nonresponse issues. To the extent that students at noncooperating base year schools may have differed from students at cooperating schools, student level bias is introduced that persists through subsequent waves of observation. Base year school nonresponse is documented and analyzed in the *NELS:88 Base Year Sample Design Report*. While the 1990 and 1992 samples were student-driven, school cooperation was sought as a cost-efficient means of surveying students, and as a source of contextual data about students' educational situations. Between 98 and 99 percent of first and second follow-up schools participated in the study.

Item nonresponse rates and patterns are documented in the various NELS:88 user's manuals. In general, missing data have not been imputed in the NELS:88 dataset. Although item response rates in NELS:88 are generally high, item nonresponse propensities vary with student characteristics (e.g., race, gender, test quartile), and hence may be a source of bias. In turn, this suggests that missing data imputation might be of value for NELS:88, a potential benefit that should be weighed against the cost of such a procedure when multiplied across the large numbers of questionnaire items in the dataset.

Finally, our discussion has not dealt with the important consideration of sampling error. Design effects for NELS:88 are documented in the various user's manuals. In this respect, dispersion of the student sample after eighth grade has been both a blessing and a curse for NELS:88. The high costs of following dispersed students required that we subsample students in the first follow-up; subsampling increases design effects. At the same time, the general tendency in a longitudinal study is for design effects to decrease over time, as dispersion reduces the original clustering.

Bibliography

Andersen, R., Kasper, J., and Frankel, M.R. (1979). ***Total Survey Error***. San Francisco: Jossey-Bass.

Babbie, E.R. (1973). ***Survey Research Methods***. Belmont, CA: Wadsworth.

Bradburn, N.M., Rips, L.J., and Shevell, S.K. (1987). "Answering Autobiographical Questions: The impact of Memory and Inference on Surveys." ***Science***, 239:157-161.

CBO (Congress of the United States, Congressional Budget Office). (1990). ***Services of Support for Adolescent Mothers***. Washington, D.C.: Author.

Duncan, G.J., Hill, M.S., and Hoffman, S.D. (1988). "Welfare Dependence Within and Across Generations." ***Science***. 239:467-471.

Fetters, W.B, Stowe, P., and Owings, J.A. (1984). ***Quality of Responses of High School Students to Questionnaire Items***. Washington, D.C.: National Center for Education Statistics.

Groves, R.M. (1989). ***Survey Errors and Survey Costs***. New York: Wiley.

Ingels, S.J., et al. ***NELS:88 Data File User's Manuals***. Base Year: Student, Parent, Teacher, School (NCES, 1990); First Follow Up: Student, Dropout, Teacher, School (NCES 1992); Second Follow-Up: Student, Dropout, Parent, Teacher, School, Transcript (NCES 1994).

Kasprzyk, D., Duncan, G., Kalton, G., & Singh, M.P., eds. (1989). ***Panel Surveys***. New York: Wiley.

Kaufman, P., Rasinski, K.A., Lee, R., and West, J. (1991). ***Quality of Responses of Eighth-Grade Students in NELS:88***. Washington, D.C.: National Center for Education Statistics.

Kish, L. (1965). ***Survey Sampling***. New York: Wiley.

Kish, L. (1987). ***Statistical Design for Research***. New York: Wiley.

Lee, V.E., and Loeb, S. (1994). "Where Do Head Start Attendees End Up? One Reason Why Preschool Effects Are Not Sustained." Paper presented to the American Educational Research Association, Annual Meeting, New Orleans, Louisiana, April 7, 1994.

McGrew, K.S., Thurlow, M.L., and Spiegel, A.N. (1993). "An Investigation of the Exclusion of Students With Disabilities in National Data Collection Programs." ***Educational Evaluation and Policy Analysis***. vol.15

Milne, A., Myers, D., Rosenthal, A., and Ginsburg, A. (1986). "Single Parents, Working Mothers, and the Educational Achievement of School Children." ***Sociology of Education***, vol. 59.

Pearson, R. W. (1989). "The Advantages and Disadvantages of Longitudinal Surveys." ***Research in the Sociology of Education and Socialization***. vol. 8.

Rock, D.A. and Pollack, J.M. (1991). ***Psychometric Report for the NELS:88 Base Year Test Battery***. Washington, D.C.: National Center for Education Statistics.

Rosenzweig, M.R., and Wolpin, K.I. (1993). "Maternal Expectations and Ex Post Rationalizations: The Usefulness of Survey Information on the Wantedness of Children." ***Journal of Human Resources***, vol. XXVIII.

Spencer, B.D., Frankel, M.R., Ingels, S.J., Rasinski, K.A., and Tourangeau, R. (1990). ***NELS:88 Base Year Sample Design Report***. Washington, D.C.: National Center for Education Statistics.

Spencer, B.D. (1991). "Eligibility/Exclusion Issues in the 1990 Trial State Assessment." In G. Bohrnstedt et al., ***Assessing Student Achievement in the States: Background Studies for the Evaluation of the NAEP Trial State Assessment***. National Academy of Education. Stanford, California: Stanford University.

Spencer, B.D., and Foran, W. (1991). "Sampling Probabilities for Aggregations, With Application to NELS:88 and Other Education Longitudinal Surveys." ***Journal of Educational Statistics***, 16(1).

END NOTES

1An alternative longitudinal design for studying students and schools might take the institutional level as the panel, perhaps freshening the school sample over time, repeatedly drawing cross-sectional samples of students within the panel of schools. Rather than collecting data from the same schools, but different student samples in different years, NELS:88 collects data from the same students.

2Longitudinal studies are prospective, in that they record new events for *individual* units under observation more or less as they occur. A repeated cross-sectional study can also be prospective, and measure change over time at the *group* level. A single (not repeated) cross-sectional study can measure change in *individual* units over time by assuming a retrospective focus--for example, by relying on individual memories to reconstruct an historical record of events and statuses. Retrospective data may also be elicited by longitudinal studies (see end note 10, below). While retrospective studies are appropriate for many purposes, when used for other purposes there may be significant reliability problems. For a concise summary of issues concerning the reliability of retrospective reports, see Bradburn, Rips and Shevell (1987). For a useful comparison of prospective and retrospective studies, see Kish (1987) pp. 178-181. For a recent example (*ex post facto* reports of wantedness of children) of an analysis of the degree to which retrospectively-obtained survey data provide unbiased estimates, see Rosenzweig and Wolpin (1993).

3We take a representative sample to be a probability sample drawn, with known selection probabilities for sample units, from the target population.

4Groves (1989, chapter 3) provides a useful discussion of coverage error and its consequences.

5There are many ways to characterize sample designs that measure change over time, and the term "longitudinal" has both strict and looser usages. Kish (1987, Chapter 6) presents a useful typology of designs for covering time spans across populations, and Babbie (1973, pp.62-66) some standard terminology.

6For individuals interested in conducting trend analyses of NLS-72, HS&B and NELS:88 data, further information on content and design similarities and differences between these three studies is presented in the second follow-up student component data file user's manual. Comparison of sophomore cohort dropouts across HS&B and NELS:88 is discussed in the dropout component user's manual, while high school transcript comparisons (HS&B, NAEP 1987, NAEP 1990, NELS:88) are discussed in the transcript user's manual.

7For each cohort, timing of the last follow-up assumes that the tentatively scheduled date for the fourth follow-up -- 1997 -- will hold. Consideration is being given to moving the data collection to 1998.

8Repeated cross-sections compound sampling error. This is the case because a repeated cross-section is drawn two or more times; change measurement must contend with the fact that differences in multiple sample means will in part be a function of the sampling errors associated with each independent sample. In contrast, a longitudinal sample is drawn but once. However, for a freshened cohort study such as NELS:88, some sampling error may be associated with the freshening process.

9The 1988-90 dropout rate for the expanded (eligible + ineligible) NELS:88 eighth grade cohort was 6.8 percent for 1988-90. Excluding students who dropped out between 1988 and 1990 (or left the country), the dropout rate between 1990 and 1992 was 7.6 percent. However, the proportion of 1988 eighth graders who were dropouts in the spring of 1992 was 11.6 percent. (Of course, the number of sample members experiencing brief duration dropout spells or dropout *events* is even further undercounted by virtue of using a cohort status [spring to spring across two years] measurement.)

10On NLSY (the BLS National Longitudinal Survey of Youth which began in 1979), see CBO, 1990. On PSID (Panel Study of Income Dynamics, a nationally-representative sample of families, begun in 1968) results, see Duncan, Hill, and Hoffman, 1988.

11Researchers (see Lee and Loeb, 1994) have used the response to this retrospective item in conjunction with NELS:88 measures of school quality to inquire into whether Head Start participants are more likely than their peers to attend lower quality elementary/middle schools, a possibility that could in part explain why academic gains from Head Start may fade out over time. This example--a question item that looks backward nine years--is an extreme case of a typical phenomenon--retrospective reporting in a prospective longitudinal study. (For an equally extreme example from HS&B, see Milne et al. 1986). More typical examples of retrospective reporting are (in NELS:88) inquiries about prior year experiences and behavior (for

BEST COPY AVAILABLE

example, given the two year data collection interval, NELS:88 elicited information from sophomores about ninth grade events and from seniors about eleventh grade events). In the NELS:88 third follow-up (1994), complete 1992-94 event histories for employment and postsecondary enrollment are being obtained retrospectively, as has been done in NLS-72, HS&B, and (on an annual basis) NLSY. Retrospective reporting frees panel studies from the limitation of their fixed time intervals of data collection. Continuous-time stochastic models (such as "hazard" or "event history" analysis) require complete information so that changes in states can be considered in relation to their precise timing. Retrospective accumulation of event histories is viable in a panel design only insofar as the recall period and the character and salience of the events assort to the degree needed to provide accurate data.

12There are plausible alternatives to this approach, based on drawing the initial (pre-high school) sample with the requirements of high school student and school representativeness in mind. For example, an integrated eighth-tenth grade sample might take into account feeder paths at this transition point, or select tenth grade schools first and eighth grade feeders contingently, or might exhaustively select schools within pre-specified districts or delimited portions of districts to maximize the overlap of "origin" and "destination" schools.

13For further details of school-level exclusion, see Spencer, Frankel, Ingels, Rasinski, & Tourangeau, 1990, p.10.

14Recent investigations of the extent to which students with disabilities are allowed to participate in major national data collection programs suggest that 40-50 percent of students with disabilities are typically excluded from major assessments, though students with disabilities are included to a greater degree in data collections that do not require the completion of cognitive tests (McGrew, Thurlow, & Spiegel, 1993). Additional numbers of students are excluded from assessments or other state and national education data collection programs owing to language barriers to participation. For a parallel discussion based on the NAEP trial state assessments, see Spencer in Bohrnstedt, ed., 1991.

15According to Office of Special Education figures reported in the *Digest of Education Statistics, 1992*, Table 51, 5.5 percent of special education students receive services in separate schools or residential facilities, while .8 percent are in a homebound or hospital environment. Not all of these individuals are in graded programs. Separate facilities tend in particular to be available for comparatively rare populations such as individuals with severe visual or hearing impairments, and for emotionally disturbed students whose presence might impede regular classroom activities. Most students who are doubly physically disabled by being both deaf and blind are educated in special facilities.

16Some of these freshened students were dropped in the subsampling process either because they themselves were not included in the subsample or because the base year student to whom they were linked was not included. Some 1,043 students selected through the freshening procedure remained in the final first follow-up sample. In the second follow-up (1992), 244 students were added through freshening.

17See Kish (1965) for a discussion of the half-open interval procedure that underpins this approach.

18A strategy for estimating a school's selection probabilities under these circumstances is sketched in Spencer and Foran, 1991.

19Again, however, while weights can compensate for nonresponse by correcting errors in the population estimates for particular subgroups, they do not correct nonresponse *bias* within subgroups. For example, weighting can adjust for the fact that male eighth graders responded to NELS:88 at a lower rate than did their female classmates, but do not address bias that may be present if male responders and nonresponders differed in the very characteristics inquired into by the base year student questionnaire.

20Discussions of longitudinal conditioning or panel effects (also known as "time in sample bias" or "panel conditioning")--for example, whether strong effects potentially exist or could affect data quality--may be found in Kasprzyk, D., Duncan, G., Kalton, G., & Singh, M.P., eds. *Panel Surveys*, 1989 (New York: Wiley). See especially contributions by B. Bailar; D. Cantor; D. Holt; A. Silberstein and C. Jacobs; L. Corder and D. Horvitz; and J. Waterton and D. Lievesley.